Serinaldi F, Lombardo F. BetaBit: A fast generator of autocorrelated binary processes for geophysical research. *EPL (Europhysics Letters)* 2017, 118(3), 30007.

DOI link to article:

https://doi.org/10.1209/0295-5075/118/30007

Date deposited:

17/07/2017

Embargo release date:

01 May 2018

BetaBit: A fast generator of autocorrelated binary processes for geophysical research

Francesco Serinaldi[1,2] and Federico Lombardo[3]

[1] *School of Civil Engineering and Geosciences, Newcastle University - Newcastle Upon Tyne, NE1 7RU, UK*
[2] *Willis Research Network - 51 Lime St., London, EC3M 7DQ, UK*
[3] *Dipartimento di Ingegneria, Università degli Studi Roma Tre - Via Vito Volterra 62, 00146 Rome, Italy*

PACS `02.50.-r` – Probability theory, stochastic processes, and statistics
PACS `02.70.Uu` – Applications of Monte Carlo methods
PACS `05.45.Tp` – Time series analysis

Abstract – We introduce a fast and efficient non-iterative algorithm, called BetaBit, to simulate autocorrelated binary processes describing the occurrence of natural hazards, system failures, and other physical and geophysical phenomena characterized by persistence, temporal clustering, and low rate of occurrence. BetaBit overcomes the simulation constraints posed by the discrete nature of the marginal distributions of binary processes by using the link existing between the correlation coefficients of this process and those of the standard Gaussian processes. The performance of BetaBit is tested on binary signals with power-law and exponentially decaying autocorrelation functions (ACFs) corresponding to Hurst-Kolmogorov and Markov processes. An application to real world sequences describing rainfall intermittency and the occurrence of strong positive phases of the North Atlantic Oscillation (NAO) index shows that BetaBit can also simulate surrogate data preserving the empirical ACF as well as signals with autoregressive moving average (ARMA) dependence structures. Extensions to cyclo-stationary processes accounting for seasonal fluctuations are also discussed.

Introduction. – Binary processes are apparently ubiquitous in physics and many other fields because every sequence of data describing a phenomenon under study can be dichotomized based upon some operational rule identifying the occurrence or non occurrence of a property of interest. Hydro-meteorological and geophysical examples include the occurrence of events or hazards such as rainfall [1], floods [2, 3], droughts [4], wind storms [5], and earthquakes [6], among others. Since these natural processes exhibit a random behavior, they can effectively be studied by statistical sampling techniques generating random sequences of real numbers that can provide insight into what constitutes "typical" behavior of real data obtained from a random experiment [7]. The availability of computers have supported the diffusion of these techniques and their application in an uncountable number of practical applications for the past seventy years [8].

Several algorithms have been developed to produce computer-generated sequences of numbers that closely resemble the samples of independent and identically distributed (iid) random variables [9] and stochastic processes with prescribed dependence structures [10]. A general approach producing sequences of real numbers with any arbitrary mean and autocorrelation function (ACF) - if it is mathematically feasible - relies on the application of a linear filter to a Gaussian white noise [11]. However, such an approach, often called the convolution method, cannot produce binary random numbers, because the output of a linear filter is a non-binary sequence even if the input is binary [12]. Since the problem of generating correlated binary sequences with specified mean is a key issue in a variety of applications in different fields, several techniques have been proposed to solve this problem. However, most of the existing methods have serious restrictions on the class of autocorrelation functions that can be effectively modeled [12–15].

Serinaldi and Lombardo [16] showed that classical spectral techniques can effectively be used if one focuses on the parent continuous process of beta distributed transition probabilities rather than on the target binary process. This change of paradigm yields a simulation procedure effectively embedding a spectrum-based iterative amplitude adjusted Fourier transform (IAAFT) method [17–19] devised for continuous processes, and allows the simulation of binary processes with prescribed dependence structure and surrogate data reproducing the empirical ACF of the observed sequences.

In this study, we describe an alternative non-iterative approach, called BetaBit, that effectively exploits the existing relationship between the pairwise correlation coefficients of bivariate Gaussian distributions and those of bivariate binary processes. We compare a procedure based on the numerical inversion of this relationship to our alternative version relying on closed form approximation formulas based on beta distribution. The performance of BetaBit is tested by simulating binary signals with power-law and exponentially decaying ACFs. The analysis and modeling of real world binary sequences describing rainfall intermittency and the occurrence of strong positive phases of the NAO index show that BetaBit can also generate surrogate data preserving the empirical ACF as well as signals with autoregressive moving average (ARMA) dependence structures under the condition of positive definiteness.

BetaBit algorithm. – We aim to generate a correlated sequence of random numbers $\{x_j\}_{j\in\mathbb{N}}$, for simplicity $\{x\}$, taking values 1 and 0 with probability p and $1-p$, respectively. The underlying discrete-time stochastic process $X = \{X_j\}_{j\in\mathbb{N}}$ with state space $\{0,1\}$, where j $(=0,1,2,\ldots)$ denotes discrete time, is specified in terms of its mean $\mu_X = \mathrm{E}[X] = p$ and autocovariance function (ACVF)

$$c_X(\tau) = \mathrm{E}\left[X_j X_{j+\tau}\right] - \mu_X^2 = \sigma_X^2 \rho_X(\tau), \tag{1}$$

where $\mathrm{E}[\cdot]$ denotes expectation (ensemble average), τ is the time lag, $\sigma_X^2 = \mathrm{Var}[X] = p(1-p)$ and $\rho_X(\tau) = \mathrm{Corr}[X_j, X_{j+\tau}]$ are the variance and autocorrelation function (ACF) of X, respectively. The simulation of correlated sequences for continuous processes can be performed by the convolution of a generic sequence of uncorrelated values $\{\varepsilon\}$ [20]. However, the direct application of the convolution method to $\{\varepsilon\}$ cannot produce correlated binary random numbers $\{x\}$, because the output of a linear filter is a non-binary sequence even if the input is binary [12].

This problem can be overcome by considering the generation of autocorrelated binary sequences as a special case of simulation of cross-correlated binary vectors [21–23]. Ideally, the most straightforward way to simulate such a binary-valued stochastic process, X, could be that of generating a sequence of n random numbers $\{y_j\}_{j=0}^{n-1}$ for an auxiliary process Y with the desired ACF (*e.g.*, exponential or power-law) and standard Gaussian marginal distribution Φ, and then transform the marginal distributions into Bernoulli marginals by

$$x_j = \begin{cases} 1 & \text{if} \quad y_j < \Phi^{-1}(p) \\ 0 & \text{otherwise} \end{cases} . \tag{2}$$

where Φ^{-1} denotes the inverse of Φ, *i.e.* the quantile function. However, this dichotomization does not preserve linear correlation. Nevertheless, it can be shown that $\rho_X(\tau)$ is related to the correlation of the process Y, $\rho_Y(\tau) = \mathrm{Corr}[Y_j, Y_{j+\tau}]$, by the relationship [21]

$$\rho_X(\tau) = \frac{\Phi_2(\Phi^{-1}(p), \Phi^{-1}(p); \rho_Y(\tau)) - p^2}{p(1-p)} = \psi(\rho_Y(\tau), p), \tag{3}$$

where Φ_2 denotes the bivariate Gaussian cumulative distribution function (cdf) of $(Y_j, Y_{j+\tau})$. Therefore, to simulate X with prescribed rate of occurrence, p, and ACF, $\rho_X(\tau)$, it is sufficient to simulate Y with corresponding $\rho_Y(\tau)$ values such that $\rho_X - \psi(\rho_Y, p) = 0$ for any lag τ. Since $\psi(\rho_Y, p)$ is continuous and strictly increasing and it ranges in the interval $[-1,1]$, the solution of eq. 3 is unique and can be found by root-finding methods [23, 24] or polynomial approximations [25].

In this study, we analyze the relationship between $\rho_X(\tau)$ and $\rho_Y(\tau)$ in order to find a functional relationship $\rho_Y = \psi^{-1}(\rho_X)$ yielding a suitably inflated ACF for the sequences $\{y\}$, so that the resulting sequences $\{x\}$ have the desired ACF after dichotomization (eq. 2), without resorting to root-finding procedures. In fact, minimizing the use of numerical solvers can be an advantage for large scale simulation studies, as it is commonly the case with geophysical problems.

Figure 1(a) shows the relationships between ρ_X and ρ_Y for varying p. For $p = 0.5$ (*i.e.*, $\Phi^{-1}(p) = 0$), $\rho_Y = \sin(2\pi\rho_X)$ [23] covering the full range of possible values $[-1,1]$, meaning that the joint distribution can describe both positive and negative linear correlation. As p decreases, the range of admissible negative values of ρ_Y tends to zero according to the theoretical lower bound $\max\{-p/(1-p), -(1-p)/p\}$ [21, 26]. Therefore, for a given p, it is not possible to simulate binary antipersistent processes characterized by negative ACF values below such a bound, regardless of the correlation of the auxiliary bivariate Gaussian process. This problem is not as serious as it looks for geophysical applications. In this context, we are usually interested in persistent occurrence (binary) processes with positive ACF, overdispersion, and clustering [1–3, 5, 27–29].

Since ρ_Y and ρ_X are linked by a monotonic nonlinear relationship (eq. 3), and $\rho_X \in [0,1]$ for $\rho_Y \in [0,1]$ (see eq. 3 and fig. 1(a)), a suitable candidate function describing such a relationship is a beta cdf G with parameters α_p and β_p depending on p

$$\begin{aligned} \rho_Y &= G(\rho_X'; \alpha_p(p), \beta_p(p)) \\ &= \frac{1}{\mathrm{B}(\alpha_p, \beta_p)} \int_0^{\rho_X'} s^{\alpha_p - 1}(1-s)^{\beta_p - 1}\mathrm{d}s. \end{aligned} \tag{4}$$

where $\rho_X' = 2/\pi \sin^{-1}((2/\pi \sin^{-1}(\rho_X^{0.25}))^{0.25})$. Notice that G in eq. 4 does not have a probabilistic interpretation,

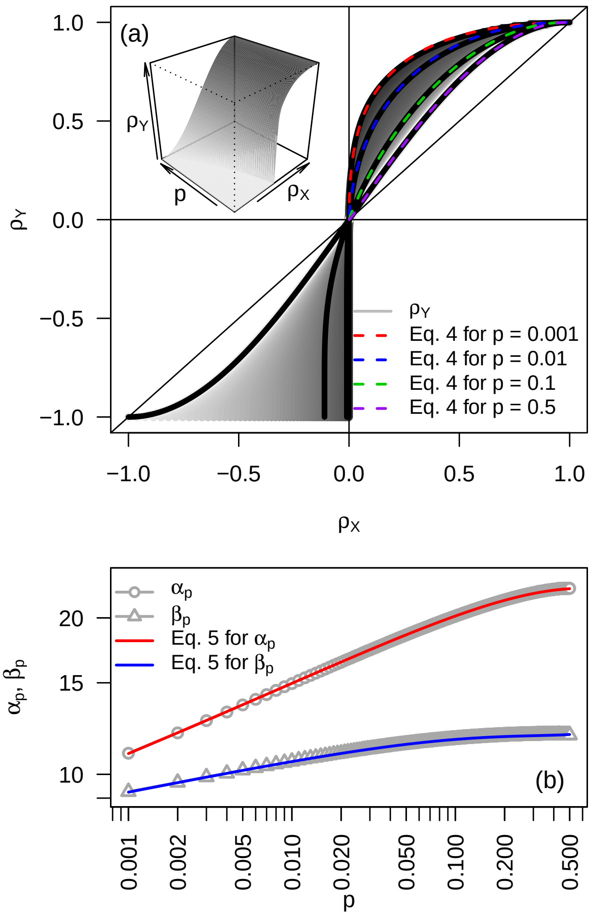

Fig. 1: (a) Law of variation of ρ_Y as a function of p and ρ_X (inset figure shows the 3-dimensional surface). Curves describing the approximation formula in eq. 4 for $p \in \{0.001, 0.01, 0.1, 0.5\}$ are reported to highlight the agreement with exact values. (b) Laws of variation of α_p and β_p as functions of p. Curves corresponding to the relationships in eq. 5 are also shown.

but it is simply used as a generic monotonic link function $f : [0,1] \to [0,1]$ to describe the relationship between ρ_X and ρ_Y. Some examples are highlighted in fig. 1(a). The approximation in eq. 4 yields errors smaller than 0.02 for $\rho_Y \in [0, 0.004]$ and 0.01 for $\rho_Y \in (0.004, 1]$, thus implying small overestimation of very low correlation values. To understand how the shape of this relationship evolves with the rate of occurrence p, we have estimated the parameters α_p and β_p for each set of pairs (ρ_X, ρ_Y) corresponding to p ranging from 0.001 to 1 by steps of 0.001. Results are summarized in fig. 1(b). The relationship between α_p, β_p and p is well approximated by generalized power law functions

$$\begin{cases} \alpha_p = 2.281 + 27.541 p^{0.167+0.07p} e^{-0.313p} \\ \beta_p = 0.11 + 14.129 p^{0.063+0.177p} e^{-0.147p} \end{cases} . \qquad (5)$$

To summarize, the implementation steps of the BetaBit algorithm to generate a correlated sequence $\{x\}$ are: (i)

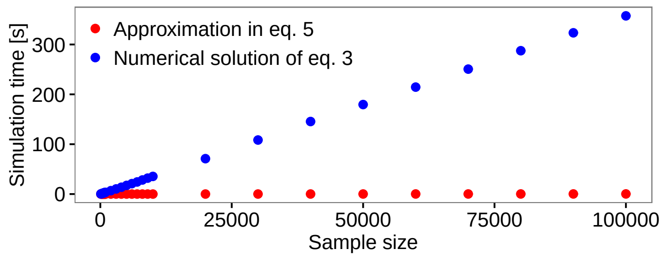

Fig. 2: Relationship between simulation time and and sample size for the algorithms based on (i) the approximated solution in eqs. 4 and 5, and (ii) the numerical inversion of eq. 3. Values refer to simulation times of binary signals with power-law decaying ACF.

compute α_p and β_p from eq. 5 based on the desired value of p; (ii) use eq. 4 to inflate the terms of the ACF of the auxiliary process Y, $\rho_Y(\tau)$, in order to obtain the target process X with the desired ACF, $\rho_X(\tau)$; (iii) generate a standard Gaussian time series $\{y\}$ with the inflated ACF by algorithms allowing the explicit use of the ACF in the simulation process, such as the method proposed by Davies and Harte [30] and used in this study; finally (iv) apply the dichotomization in eq. 2.

The great advantage of using eqs. 4 and 5 instead of numerically inverting eq. 3 is illustrated in fig. 2, which shows how the simulation time of a binary time series with power-law decaying ACF (discussed in the next section) increases with the sample size n. Both approaches require $\mathcal{O}(n)$ operations; however the growth rate is $\approx 3.6 \cdot 10^{-3}$ s/realization for the numerical solution, and $\approx 2.8 \cdot 10^{-6}$ s/realization for the approximated solution, meaning that simulating a series of size 10^5 requires ≈ 360 s and ≈ 0.28 s, respectively (see fig. 2).

Simulation of binary signals with exponentially and power-law decaying ACF, and extension to cyclo-stationary processes. – The performance of BetaBit is tested by generating binary sequences with ACF corresponding to two widely used stationary processes, *i.e.* the Hurst-Kolmogorov (HK) and the Markov process. The former, also known as fractional Gaussian noise, is characterized by the following ACF

$$\rho_X(\tau) = \frac{1}{2}\left(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H}\right), \qquad (6)$$

which exhibits a power-law decay $\rho_X(\tau) \propto |\tau|^{2H-2}$. For $0.5 < H < 1$ the process is positively correlated and exhibits long-range dependence, while it reduces to white noise for $H = 0.5$. As a second example, we consider a process with short-range Markovian dependence, which is characterized by exponentially decaying ACF of the form

$$\rho_X(\tau) = \exp(-\gamma|\tau|) = \rho_1^{|\tau|}, \qquad (7)$$

where $1/\gamma$ is the correlation radius and $\rho_1 = \exp(-\gamma)$ is the lag-one autocorrelation coefficient. Simulations are

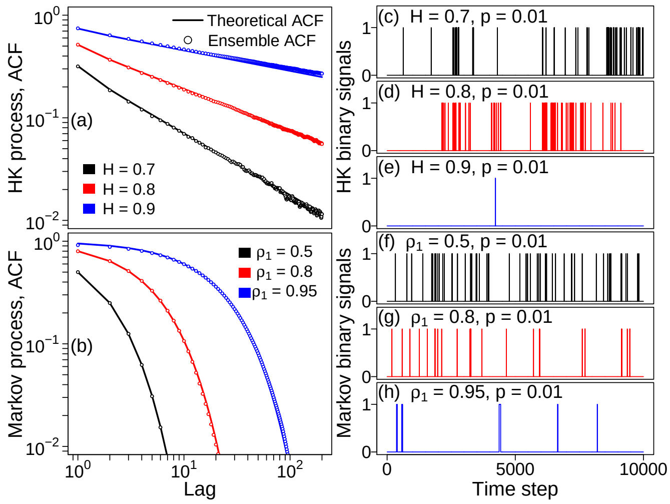

Fig. 3: ACFs (a-b) and sample signals (c-h) corresponding to HK and Markov processes with three different values of the characteristic parameters (H and ρ_1), and $p = 0.01$. Simulation relies on ρ_Y values computed by inverting eq. 3 via a numerical root-finding procedure. Panels (a-b) show the analytical ACFs (—) along with the empirical ACFs of the simulated binary sequences $\{x\}$ ($\bullet$) for each of the parameter values reported in the legends. Panels (c-e) and (f-h) show synthetic sequences corresponding to ACFs reported in (a) and (b), respectively. The simulated sequences exhibit an increasing clustering effect related to the increasing strength of the autocorrelation (*viz.* parameter values).

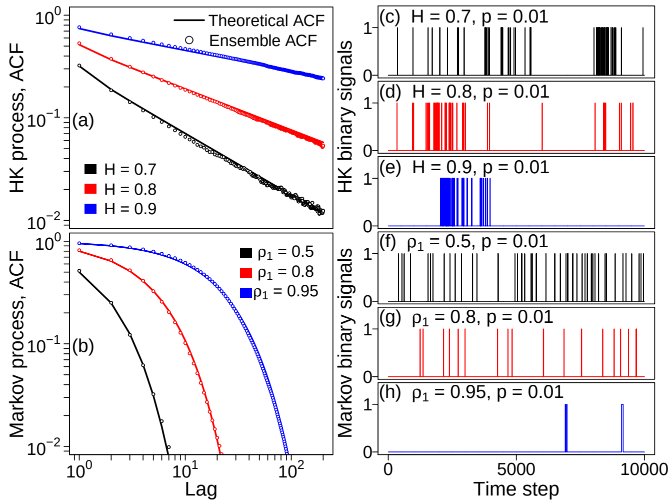

Fig. 4: As fig. 3 but for ρ_Y approximated by formulas in eqs. 4 and 5; the same caption and interpretation apply.

performed for $H \in \{0.7, 0.8, 0.9\}$, $\rho_1 \in \{0.5, 0.8, 0.95\}$, $p \in \{0.01, 0.05, 0.1\}$. The values of p may mimic the rate of occurrence of rare events such as storms, floods, earthquakes and other geophysical hazards.

Results for $p = 0.01$ are illustrated in figs. 3 and 4, which refer to ρ_Y values computed by inverting eq. 3 via a numerical root-finding procedure and by approximation formulas in eqs. 4 and 5, respectively. Figures 3(a)-(b) and 4(a)-(b) compare each theoretical ACF with the empirical ensemble counterpart for HK (figs. 3(a) and 4(a)) and Markov (figs. 3(b) and 4(b)) processes with the set of parameters mentioned above. Ensemble ACFs are estimated from 5000 sequences of size 2000. The agreement between theoretical and empirical ACFs denotes the effectiveness of the proposed approach to simulate binary signals with power-law and exponentially decaying ACFs, as well as with prescribed mean and variance. Therefore, BetaBit effectively generates correlated binary sequences of random numbers by means of a closed-form algorithm, which is a faster alternative to the numerical simulation procedures existing in the literature.

For each model, figs. 3(c)-(h) and 4(c)-(h) show some examples of synthetic sequences of length 10000. In all cases, values 0 and 1 tend to cluster more and more as the degree or extent of autocorrelation increase. This clustering behavior is in agreement with previous theoretical and empirical findings resulting from the study of extreme values of simulated and observed processes taking real values [2, 3, 27, 28, 31–35].

Even though our algorithm is devised to simulate binary sequences with ACF corresponding to that of stationary processes, seasonal patterns characterizing the occurrence of many geophysical variables, such as rainfall, stream flow and sea waves [3, 36, 37], can be incorporated in BetaBit in order to obtain cyclo-stationary binary signals. This can be done by stratifying data on a monthly or seasonal basis, then estimating the empirical ACF and/or the required parameters of the theoretical models for each subset, and finally merging the simulated series corresponding to each month or season. Alternatively, if the seasonal variation is mainly related to a cyclic behavior of the rate of occurrence (*e.g.*, less rainfall events in summer than in winter), then seasonal fluctuations of p can be introduced by means of a periodic function reproducing the different rate of occurrence corresponding to each season in the same spirit of other generators of point processes, such as the integrate-and-fire method [38]. Both methods yield non-homogenous persistent point processes. The former approach allows for the seasonal variation of both p and ACF parameters. The latter implies seasonal variation of p but constant ACF parameters, and its rationale is similar to that of ARMA modeling of deseasonalized time series.

For the sake of illustration, we show an example of application of the second method as a proof of concept. Fig. 5 shows the average ACFs (over 5000 sequences of size $10 \cdot 365$) and some sequences generated by using a time-varying p with average value $\bar{p} = 0.2$ and sinusoidal fluctuation $p(t) = 0.19 \sin(2\pi t/365) + 0.2$ with period equal to 365. The sequence of values yielded by function $p(t)$ is used in the dichotomization in eq. 2, where $\{y\}$ are sequences corresponding to Markov processes with $\rho_1 \in \{0.5, 0.8, 0.95\}$.

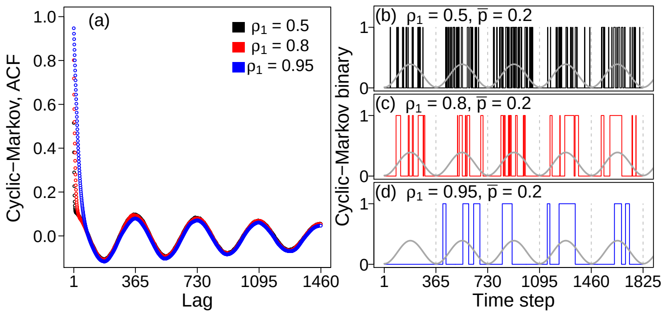

Fig. 5: Average ACFs (a) and sample signals (b-d) corresponding to cyclo-stationary processes with underlying Markov dependence structure with three different values of ρ_1, and $p(t) = 0.19 \sin(2\pi t/365) + 0.2$ (continuous gray lines in panels (b-d)). The simulated sequences exhibit an increasing clustering effect around the periods with high p values (which is related to the increasing strength of the autocorrelation), while preserving systematic lack of events when $p \approx 0$.

According to this model structure, each simulated binary sequence of length 10·365 mimics ten years of daily records with a hypothetical summer/winter alternation in terms of rate of occurrence p. The simulated time series show that the events tend to cluster around the time steps t where p reaches the maximum value, while few or no events occur when p is close to zero. As ρ_1 increases, events tend to cluster more and more, generating consecutive (persistent) sequences of occurrences in the periods with high p values, while preserving systematic lack of occurrences when $p \approx 0$.

These remarks further highlight the efficiency and flexibility of BetaBit. However, since these advanced aspects require deeper investigation, the examples of applications to geophysical variables shown in the next section refer to homogenous processes, according to the approach based on data stratification.

Geophysical applications. – As mentioned in the introduction, binary sequences are very common in geophysics as they naturally rise when one focuses on the occurrence/non-occurrence or presence/absence of a given event and/or characteristic. For the sake of illustration, we consider both rainfall intermittency, *i.e.* the alternation of wet and dry periods, and the occurrence of positive phases of the NAO index exceeding the 80% threshold. As shown below, this analysis allows for testing (i) the performance of BetaBit against an additional parametric dependence structure, namely ARMA ACFs (with positive ACF terms), and (ii) the capability of the proposed methodology to generate so-called surrogate sequences preserving on average the observed $\hat{p}$ and empirical autocorrelation function. Surrogate data are useful for exploratory analysis to check, for instance, nonlinearity [17–19, 25] or clustering and departures from Poissonian behavior (or from other benchmark processes) in the context of occurrence (binary) data such as timings of climatic extreme events or neuronal spikes in neurosciences [23, 39]. On the other hand, parametric models can be used for prediction, sensitivity analysis, or as a part for more general models [40].

The dependence structure of the rainfall occurrence process appears to be non-Markovian [1], and the reproduction of the observed rate of occurrence, p, is fundamental to set up, for instance, occurrence modules of rainfall models used in hydrological studies [40]. We consider a rainfall time series recorded in Gubbio (central Italy) at 30-minute temporal resolution from 1995 to 2001 extracted from a wider data set of 35 time series previously studied [41, 42]. In order to have a homogenous sequence free of seasonal fluctuations, we focused on October data (similar results can be obtained for the other months). Figures 6(a)-(b) show the time series of rainfall depth and rainfall occurrence. The estimated rate of occurrence is $\hat{p} = 0.05$. The empirical ACF was computed on the merged sample comprising October records for the years 1995-2001, taking care of removing the cross-products of lagged observations, x_j and $x_{j+\tau}$, not belonging to the same year [43]. The occurrence process was modeled by an ARMA(2,4) dependence structure resulting from a model selection procedure based on the Akaike information criterion [44]. ARMA binary sequences are compared with surrogate series preserving the empirical ACF. The two example time series in figs. 6(c)-(d) show that both ARMA and surrogate binary sequences resemble the observed occurrences quite closely, mimicking the typical clustering behavior of rainfall events. Figure 6(e) compares the empirical ACF of the observed process with the average ACF computed from the ACFs of 1000 simulated sequences, with length equal to the one of the observed time series. The 95% confidence bands of the ACF show that the observed ACF is close to the ARMA structure for the first 25 lags, while the model tends to slightly underestimate the observed ACF for larger lags. However, the aim is not to suggest a highly-parameterized ARMA dependence structure for rainfall intermittency, but to show the ability of BetaBit to reproduce a variety of ACF models. The comparison between the observed ACF and the average and the 95% confidence bands of the ACFs resulting from 1000 surrogate series (fig. 6(f)) confirms that BetaBit easily generates accurate surrogate series preserving on average the observed ACF with limited variability around the expected pattern.

The study of the occurrence of the NAO strong positive phases is of interest because they tend to be associated with above-normal temperatures in the eastern United States and across northern Europe and below-normal temperatures in Greenland and often across southern Europe and the Middle East. They are also associated with above-normal precipitation over northern Europe and Scandinavia and below-normal precipitation over southern and central Europe [45]. Since this behavior is most pronounced during winter, we focused on NAO data from December to March [45, 46].

Figure 7(a) shows the first 3000 values of the winter

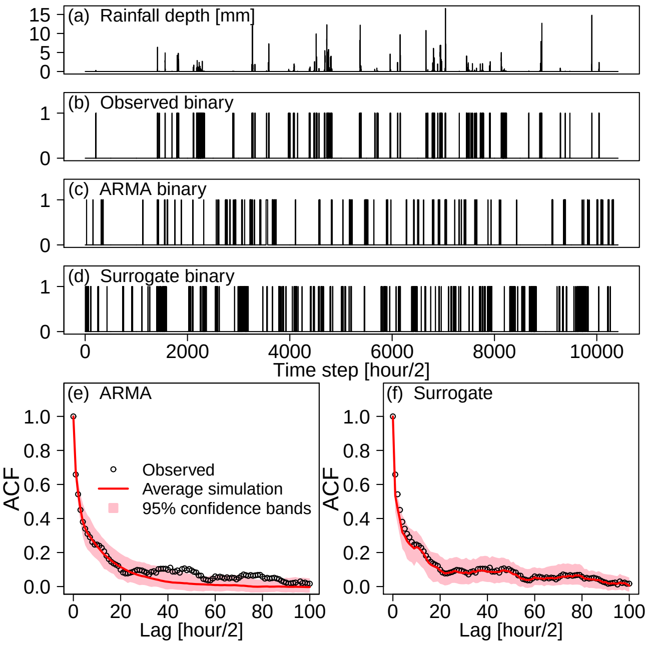

Fig. 6: (a) Time series of rainfall depth [mm] obtained by merging October data recorded in Gubbio (central Italy) at 30-minute temporal resolution from 1995 to 2001. (b) Observed occurrence process corresponding to the rainfall records in panel (a). (c-d) Typical synthetic time series, of equal length, generated by the BetaBit algorithm respectively with ARMA dependence structures, and empirical ACF (surrogate sequence). Comparison with the observed occurrence process in panel (b) shows that ARMA model and surrogate can reproduce the typical clustering behavior (also known as over-dispersion) of rainfall events. (e-f) Comparison of the ACF of the observed occurrence process and the mean ACF obtained by averaging the ACFs of 1000 synthetic signals with the same size of the observed sequence. The 95% confidence bands are also reported. Panels (e-f) show that the average ACF of ARMA sequences fits well the observed ACF, even though sampling fluctuations fall outside the confidence bands for lags larger than 25. The average ACF of surrogate series in panel (f) closely follows the empirical ACF with limited fluctuations as expected according to the definition of the surrogate.

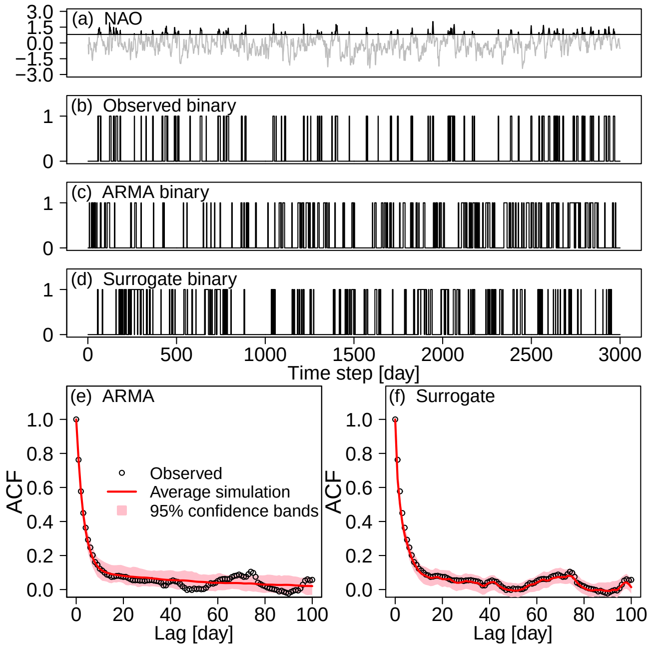

Fig. 7: As fig. 6 but for the occurrence of the NAO phases above the 80% quantile; the same caption and interpretation apply. The fluctuations of the simulated ACFs are smaller because of the larger sample size and rate of occurrence.

NAO daily records spanning the period December 1st, 1950 to March 31st, 2016, along with the 80% threshold. In this case, the selected model is an ARMA(2,2), $\hat{p} = 0.2$, and the empirical ACF was estimated by the same approach used for rainfall data. A visual comparison of the binary sequence of the over-threshold occurrences and ARMA and surrogate sequences (figs. 7(b)-(d)) confirms the ability of BetaBit to generate binary time series similar to the observed ones. The overlap of the observed ACF and the average ACFs from 1000 simulated sequences (figs. 7(e)-(f)) highlights the agreement between the dependence structures.

Conclusions. – We introduced an efficient non-iterative algorithm, called BetaBit, to generate binary processes with given mean and autocorrelation describing the occurrence of natural hazards, system failures, and other physical and geophysical phenomena showing persistence and clustering. BetaBit is a fast alternative to the algorithm proposed by Serinaldi and Lombardo [16] in the spirit of Macke *et al.* [23]. It relies on the relationship between the correlation coefficient of auxiliary bivariate Gaussian processes and that of the target bivariate binary processes. In particular, we introduce an analytical approximation of such a relationship to avoid its inversion via numerical solvers, and allow for a faster simulation. The methodology is fully general and enables the use of desired autocorrelation structure under minimal assumptions. Such a flexibility was tested by simulating sequences from models with exponentially and power-law decaying autocorrelation, and surrogate data preserving the empirical correlation structure. The former are useful for prediction, sensitivity analysis, or as a part for more general models, while the latter for exploratory purposes. We also provided a preliminary discussion of possible extension of BetaBit to simulate cyclo-stationary processes with periodic fluctuations of the rate of occurrence p.

The application to real world data describing rainfall intermittency and the occurrence of strong phases of the NAO index showed that BetaBit reproduces the typical clustering behavior exhibited by rainfall or other geophysical processes as well as extreme events such as floods, wind storms, and others characterized by low rate of occurrence. Therefore, BetaBit can effectively be used in risk analysis. For example, simulating the number of extreme events (*e.g.*, floods or storms) over a time window allows one to assess the variability of the number of failures we can expect in the design life of a system (*e.g.*, defense infrastructure). These synthetic experiments can be performed in a verification setting by starting from the rate of occurrence and dependence structure of recorded data, or for a sensitivity analysis exploring the effect of the rate of occurrence and persistence.

* * *

FS acknowledges support from Willis Research Network. FL is grateful to Silvia Ghinaglia for her continuous support and fruitful discussions. Remarks from two anonymous reviewers are gratefully acknowledged.

REFERENCES

[1] Koutsoyiannis D., *Water Resour. Res.*, **42** (2006) W01401.
[2] Bogachev M. I. and Bunde A., *Europhys. Lett.*, **97** (2012) 48011.
[3] Serinaldi F. and Kilsby C. G., *Water*, **8** (2016) 152.
[4] Coats S., Smerdon J. E., Karnauskas K. B. and Seager R., *Environ. Res. Lett.*, **11** (2016) 074025.
[5] Mailier P. J., Stephenson D. B., Ferro C. A. T. and Hodges K. I., *Mon. Weather Rev.*, **134** (2006) 2224.
[6] Ogata Y., *Pure Appl. Geophys.*, **155** (1999) 471.
[7] Papoulis A., *Probability, Random Variables, and Stochastic Processes* 3rd Edition (McGraw Hill, New York) 1991.
[8] Kroese D. P., Brereton T., Taimre T. and Botev Z. I., *Wiley Interdisciplinary Reviews: Computational Statistics*, **6** (2014) 386.
[9] Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P., *Numerical Recipes: The Art of Scientific Computing* 3rd Edition (Cambridge University Press, Cambridge) 2007.
[10] Kasdin N. J., *P. IEEE*, **83** (1995) 802.
[11] Box G. E. P., Jenkins G. M., Reinsel G. C. and Ljung G. M., *Time series analysis: forecasting and control* 5th Edition (John Wiley & Sons, Hoboken, New Jersey) 2015.
[12] Izrailev F. M., Krokhin A. A., Makarov N. M. and Usatenko O. V., *Phys. Rev. E*, **76** (2007) 027701.
[13] Boufounos P., *Generating binary processes with all-pole spectra* in proc. of *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07* Vol. 3 2007 pp. 981–984.
[14] Rojas C. R., Welsh J. S. and Goodwin G. C., *A receding horizon algorithm to generate binary signals with a prescribed autocovariance* in proc. of *2007 American Control Conference* 2007 pp. 122–127.
[15] Usatenko O. V., Melnik S. S., Apostolov S. S., Makarov N. M. and Krokhin A. A., *Phys. Rev. E*, **90** (2014) 053305.
[16] Serinaldi F. and Lombardo F., *Phys. Rev. E*, **95** (2017) 023312.
[17] Schreiber T. and Schmitz A., *Phys. Rev. Lett.*, **77** (1996) 635.
[18] Kugiumtzis D., *Phys. Rev. E*, **60** (1999) 2808.
[19] Schreiber T. and Schmitz A., *Physica D*, **142** (2000) 346.
[20] Koutsoyiannis D., *Water Resour. Res.*, **36** (2000) 1519.
[21] Emrich L. J. and Piedmonte M. R., *Am. Stat.*, **45** (1991) 302.
[22] Qaqish B. F., *Biometrika*, **90** (2003) 455.
[23] Macke J. H., Berens P., Ecker A. S., Tolias A. S. and Bethge M., *Neural Comput.*, **21** (2009) 397.
[24] Cario M. C. and Nelson B. L., *Oper. Res. Lett.*, **19** (1996) 51.
[25] Kugiumtzis D., *Phys. Rev. E*, **66** (2002) 025201.
[26] Prentice R. L., *J. Am. Stat. Assoc.*, **81** (1986) 321.
[27] Bunde A., Eichner J. F., Kantelhardt J. W. and Havlin S., *Phys. Rev. Lett.*, **94** (2005) 048701.
[28] Eichner J. F., Kantelhardt J. W., Bunde A. and Havlin S., *The statistics of return intervals, maxima, and centennial events under the influence of long-term correlations* in *In Extremis*, edited by Kropp J. and Schellnhuber H.-J., (Springer Berlin Heidelberg) 2011 pp. 2–43.
[29] Telesca L., Eisner L., Stabile T. A. and Vlček J., *Europhys. Lett.*, **116** (2016) 59002.
[30] Davies R. B. and Harte D. S., *Biometrika*, **74** (1987) 95.
[31] Eichner J. F., Kantelhardt J. W., Bunde A. and Havlin S., *Phys. Rev. E*, **73** (2006) 016130.
[32] Bogachev M. I., Eichner J. F. and Bunde A., *Phys. Rev. Lett.*, **99** (2007) 240601.
[33] Eichner J. F., Kantelhardt J. W., Bunde A. and Havlin S., *Phys. Rev. E*, **75** (2007) 011128.
[34] Bogachev M. I., Eichner J. F. and Bunde A., *Eur. Phys. J.-Spec. Top.*, **161** (2008) 181.
[35] Volpi E., Fiori A., Grimaldi S., Lombardo F. and Koutsoyiannis D., *Water Resour. Res.*, **51** (2015) 8570.
[36] Serinaldi F. and Kilsby C. G., *Physica A*, **392** (2013) 1080.
[37] Besio G., Briganti R., Romano A., Mentaschi L. and De Girolamo P., *Nat. Hazards Earth Syst. Sci.*, **17** (2017) 505.
[38] Turcott R. G. and Teich M. C., *Ann. Biomed. Eng.*, **24** (1996) 269.
[39] Donges J. F., Schleussner C., Siegmund J. F. and Donner R. V., *Eur. Phys. J. Spec. Top.*, **225** (2016) 471.
[40] Lombardo F., Volpi E., Koutsoyiannis D. and Serinaldi F., *Water Resour. Res.*, (2017) doi: 10.1002/2017WR020529.
[41] Serinaldi F., *Stoch. Env. Res. Risk A.*, **22** (2008) 671.
[42] Serinaldi F., *Stoch. Env. Res. Risk A.*, **23** (2009) 677.
[43] Papalexiou S.-M., Koutsoyiannis D. and Montanari A., *J. Hydrol.*, **411** (2011) 279.
[44] Akaike H., *IEEE T. Automa. Contr.*, **19** (1974) 716.
[45] Hurrell J. W., *Science*, **269** (1995) 676.
[46] Stephenson D. B., Pavan V. and Bojariu R., *Int. J. Climatol.*, **20** (2000) 1.